**Local Low-Rank Matrix Approximation with Preference Selection of Anchor Points**

# Abstract

矩阵分解被广泛用于个性化推荐系统，文本挖掘和计算机视觉。构造矩阵近似的一般假设是原始矩阵具有全局低秩，而Joonseok Lee等。提出许多实矩阵可能不是全局低秩，因此提出了局部低秩矩阵近似方法[11]。然而，这种矩阵近似方法仍然存在一些未解决的重要问题，例如，随机选择锚节点。在本文中，我们研究了锚节点的选择问题，以增强局部低秩矩阵近似。我们提出了一种局部低秩矩阵近似的新模型，该模型使用启发式方法选择锚点。我们的实验表明，所提出的方法优于许多最先进的推荐方法。此外，该方法可以显着提高算法效率，并且易于并行化。这些特性使其具有大规模现实世界推荐系统的潜力。

# INTRODUCTION

自20世纪90年代中期第一篇关于协同过滤的论文出现以来，推荐系统已经成为一个重要的研究领域[7,16,18]。在过去十年中，业界和学术界都做了很多工作，他们专注于开发推荐系统的新方法。基于用户兴趣和购买行为特征的个性化推荐，提供他们感兴趣的用户信息和商品。随着电子商务的不断扩大，以及商品数量和种类的快速增长，客户需要花费大量时间来寻找自己想要的东西。访问大量无关信息和产品的过程无疑将继续失去客户，并且肯定会淹没信息过载问题。为了解决这些问题，个性化的推荐系统应运而生。推荐系统引起了多学科的关注，并且已经提出了许多技术来构建推荐系统。

显然，推荐方法是推荐系统最关键的部分。可以提到四种基本的推荐方法：人口过滤，协作和基于内容的推荐，以及简化的统计方法[8]。协同过滤推荐技术是推荐系统中最成功的技术之一。它通常基于用户明确提供的项目评级。该方法推荐的产品已被其他类似用户或一组此类用户积极评估，其评级与当前用户具有最强的相关性[6]。

最成功的协同过滤方法之一是概率矩阵分解（PMF）[17]。矩阵分解的目标是将用户 - 项目评级矩阵分解为用户因子矩阵和项目因子矩阵。该矩阵分解模型与观测数量成线性比例，更重要的是，在大型，稀疏和非常不平衡的数据集上表现良好。

关于PMF的一个基本假设是原始矩阵是全局低秩矩阵，这表明假设矩阵具有低秩是合理的。然而，Joonseok Lee提出了局部低秩矩阵近似（LLORMA）[11]，假设矩阵具有局部低秩而不是全局低秩。该方法随机选择一些锚点。然后，它估计锚点的每个邻域的局部低秩矩阵近似。最后，线性组合局部矩阵模型以预测新的用户项目评级。LLORMA将观察到的矩阵视为多个矩阵的叠加。

在LLORMA中，关键步骤是为每个局部矩阵随机选择锚点。**该锚点选择方法可能导致性能差和效率低，因为锚点可能在稀疏区域或重叠区域中。**在本文中，我们研究如何在局部低秩矩阵分解模型中选择锚点，并提出一种优选的锚点选择方法，用于LLORMA的聚类（称为CLLORMA）。该方法提高了局部低秩矩阵分解的有效性和效率。**基本思想是通过聚类方法生成候选锚点**，**然后基于区域密度和锚点距离标准选择相应的锚点。首先，我们将观察到的评级矩阵分解，以获得用户（行）和项目（列）的特征。接下来，我们对用户和项目因子矩阵进行聚类，以同时获得候选锚点和距离。然后我们启发式地从这些候选锚点中选择锚点。最后，我们结合锚点周围建立的所有局部矩阵来预测原始矩阵。**实验表明，基于聚类密度和锚点距离的启发式点选择方法比局部低秩矩阵分解模型中的随机点选择方法更准确，更有效。

本文的其余部分安排如下。第2节介绍了相关工作，然后在第3节中详细介绍了所提出的聚类局部低秩矩阵近似（CLLORMA）模型。实验和分析见第4节。最后，我们在第5节中总结了这篇论文。

# RELATED WORK

推荐系统是信息过载问题的解决方案，通过利用用户项交互信息或用户和项目内容信息，帮助用户找到感兴趣的对象。多年来，推荐系统引起了如此多的关注，并且已经提出了许多技术来提供推荐服务。

根据所使用的推荐信息，我们可以将当代推荐方法大致分为三类[19]：基于用户 - 项目交互信息，基于社会关系信息[1,3,13,15,22]和基于异构信息。随着社交媒体的普及，社交推荐技术[1,3,13,14,15,22]不断涌现，主要利用用户之间丰富的社交关系，例如Twitter的following关系。随着网络信息的日益复杂，异构信息网络已成为最热门的研究课题之一。并且创建了基于异构信息网络的一系列推荐技术[4,12,19,20,23,24,25]。

但是，为了提高效率和部署，推荐系统通常使用协同过滤（CF）[5]构建。CF技术仅依赖于用户项目过去交互信息，例如用户的浏览历史或产品评级。为了建立建议，它需要收集用户项数据。有两种类型的CF模型：邻域方法和隐性特征模型。

邻域方法集中于基于用户项交互矩阵（例如，用户项评级矩阵）获得项之间或用户之间的相似性信息。面向用户的方法计算用户之间的相似性，并基于对相同项目的类似用户的评级来评估用户对项目的偏好。对称地，面向项目的方法计算项目之间的相似性，并基于由同一用户评定的类似项目的评级来评估用户对项目的偏好。

自Netflix奖竞赛[2]开始以来，潜在因子模型，即最着名的低秩矩阵分解，已经在推荐系统中显示出其有效性和效率。这些年来，许多研究人员已经被这一学科所吸引，并提出了一系列基于低阶矩阵分解的方法。Salakhutdinov等人提出了概率算法，该算法与观测数量呈线性关系，并提出了概率矩阵分解（PMF）[17]。Yehuda Koren [9]设计了一个类似SVD的低秩分解评级矩阵，并将这个基本模型称为SVD。为了获得更准确的评级预测，Koren进一步将隐式反馈整合到SVD中，并提出了一个名为SVD ++的更新版本。

Lee没有假设用户项目矩阵在全球范围内排名较低，而是在[18]假设用户项矩阵在某些行列组合附近表现为低秩矩阵，则提出局部低秩矩阵近似（LLORMA）。最近，作者进一步将LLORMA与排名损失的一般经验风险最小化相结合，并提出了一种称为局部协作排序（LCR）的模型[10]。